

Certainty and Uncertainty in Quantum Information Processing

Eleanor G. Rieffel

FX Palo Alto Laboratory
rieffel@fxpal.com

Abstract

This survey, aimed at information processing researchers, highlights intriguing but lesser known results, corrects misconceptions, and suggests research areas. Themes include: certainty in quantum algorithms; the “fewer worlds” theory of quantum mechanics; quantum learning; probability theory versus quantum mechanics.

This idiosyncratic survey delves into areas of quantum information processing of interest to researchers in fields like information retrieval, machine learning, and artificial intelligence. It overviews intriguing but lesser known results, corrects common misconceptions, and suggests research directions. Three types of applications of a quantum viewpoint on information processing are discussed: quantum algorithms and protocols; quantum proofs for classical results; the use of formalisms developed for quantum mechanics in other areas with linear algebraic or probabilistic components. This paper is not tutorial in nature; readers new to the field should read it in conjunction with a tutorial (Rieffel & Polak 2000) or book (Nielsen & Chuang 2001; Rieffel & Polak in preparation) on the subject.

A number of themes underlie this paper: certainty in quantum algorithms and quantum mechanics, including a “fewer worlds” correction to popular conceptions of the “many worlds” interpretation of quantum mechanics; relations and distinct differences between probability theory and quantum mechanics, including how entanglement differs from correlation; what is known and what remains uncertain as to the source of the power of quantum information processing. The most startling thing about quantum mechanics is not that it is probabilistic, but rather that it disobeys fundamental laws of probability theory. A common framework encompassing both probability theory and quantum mechanics throws light on many of these themes. The most technical parts of the paper establish this framework and discuss its implications.

What is and isn’t quantum information processing

Quantum information processing includes quantum computation and cryptographic and communication protocols like quantum key distribution and dense coding. Quantum computation is not synonymous with using quantum effects in computation; quantum mechanical effects are used in the processors of all state of the art (classical) computers. The distinction between classical and quantum computation is whether the information being processed is encoded in a classical or quantum way, in bits or qubits.

Certainty in quantum mechanics

Non-probabilistic quantum algorithms

Glaringly obvious - perhaps blindingly so - examples of non-probabilistic quantum algorithms exist: quantum analogs of classical non-probabilistic algorithms. Any reversible classical computation has a directly analogous quantum computation. Any classical computation has a reversible counterpart using at most $O(t^{1+\epsilon})$ time and $O(s \log t)$ space (Bennett 1989). If the initial classical algorithm is non-probabilistic, so are the analogous reversible and quantum algorithms.

More surprising perhaps is that the first truly quantum algorithms - ones that do not have classical counterparts - succeed with certainty. The quantum algorithm for Deutsch’s problem (Deutsch 1985; Deutsch & Jozsa 1992) succeeds with certainty. Grover’s search algorithm is not inherently probabilistic. His initial algorithm succeeded only with high probability (Grover 1997), but with a little cleverness Grover’s algorithm can be modified so that it is guaranteed to find an element being searched for while still preserving the quadratic speed up. (Brassard, Høyer, & Tapp 1998) suggest two approaches. In essence, the first rotates by a slightly smaller angle at each step, while the second changes only the last step to a smaller rotation. Shor’s factoring algorithm is inherently probabilistic just like many of the best classical algorithms for related problems like primality testing.

Fewer worlds theory of quantum mechanics

Many papers discuss the pros and cons of the many worlds theory. Here we mean to correct not that theory, but the

popular conception of it as “everything happens in some universe”. Popular accounts of quantum mechanics, and some scholarly articles, give the impression that quantum mechanics, at least in the many worlds interpretation, implies that everything happens in some universe. A typical quote (Deutsch 1998): “There are even universes in which a given object in our universe has no counterpart - including universes in which I was never born and you wrote this article instead.” The variety of imaginative examples suggest that anything we can conceive of, even the highly unlikely, happen, if only in a small number of universes. But much of the surprise of quantum mechanics is that certain things we thought would happen, even things we thought were sure to happen, do not happen at all.

Most startling are events that were predicted to happen with certainty by classical physics, but which in fact happen with probability 0. Thus, not only is it not true that everything we can conceive of is predicted to happen in some universe, but things we can hardly conceive of not happening do not happen, not in any universe. To emphasize this correction, I call it “the fewer worlds than we might think” interpretation of quantum mechanics, or the “fewer worlds” theory for short.

Here are a few examples. In the double slit experiment, quantum mechanics predicts that no light reaches certain spots. And indeed no light reaches those spots, even though classically we expect some photons to reach every spot. Even more striking is the GHZ experiment (Greenberger, Horne, & Zeilinger 1989; Greenberger *et al.* 1990; Pan *et al.* 2000) in which the classical prediction is that each of four things happen with equal probability and another four things never happen. Quantum mechanics predicts, and experiments confirm, that the four outcomes that are classically predicted to happen never happen (and the four classically prohibited outcomes do occur, with equal probability). As a final example, we saw that many quantum algorithms return a result with probability 1; the obvious conclusion is that the other results do not happen at all.

Uncertainty in classical physics

Both relativity and uncertainty principles exist in purely classical settings. The revolutions of the 20th century, special and general relativity and quantum mechanics, expanded on these principles. In special relativity, Einstein made Galilean relativity - the notion that the speed of an object depends on the observer and is not a property of the object itself - compatible with the notion of a constant speed of light, the same for all observers. Quantum mechanics took standard classical uncertainty principles involving waves and applied them to particles with the implication that nothing of a pure particle nature exists, in this way resolving various experimental and theoretical issues.

That a particle cannot simultaneously have both a precisely defined position and a precisely defined momentum is the startling content of Heisenberg’s uncertainty principle. This statement is less surprising when applied to a wave. Uncertainty principles for classical waves are well known. For example, consider a signal $s(t)$ with a finite mean \hat{t} and

standard deviation Δt . Similarly assume the mean $\hat{\omega}$ and standard deviation $\Delta\omega$ of $s(t)$ ’s frequency distribution can be calculated. Classical signals $s(t)$ obey the uncertainty principle $\Delta t \Delta\omega \geq 1/2$. That a signal with small standard deviation in time cannot have too small a standard deviation in its frequency spectrum is not mysterious. Details can be found in many signal processing books; (Cohen 1995) is particularly detailed and insightful.

This discussion makes no mention of measurement (though it certainly has implications for measurement). Contrary to popular belief, Heisenberg’s uncertainty principle is not about imprecision in our ability to measure (though it has implications for measurement). Just like time/frequency in the signal case, Heisenberg’s uncertainty principle says that a particle cannot have definite values for both its position and momentum. The implication is that there are no classical point particles, with position and momentum both precisely defined; there aren’t even arbitrary close approximations to such. The implications of this principle for measurement is that even in an ideal case, in which measurement of a series of particles in identical states were performed perfectly, if the standard deviation of the results for position measurements is small enough then the standard deviation of the results for momentum must be proportionally large. Initially Heisenberg and others confused two arguments, one based on the wave nature of particles, the other based on a disturbance theory of measurement. It is the former that has stood the test of time. The failure of a disturbance theory was established by the famous EPR paper (Einstein, Podolsky, & Rosen 1935) (though it took decades before a fuller understanding of the implications of the EPR paradox was achieved by Bell).

Generalized uncertainty principles exist for many other pairs of properties. For example, an uncertainty relation for polarization says that if a particle has polarization close to horizontal or vertical it cannot have polarization close to 45°. This uncertainty principle is more intuitive than that for position and momentum, but the mathematics is closely related.

Applications of a quantum viewpoint to information processing

There exist three distinct classes of applications of the viewpoint that has developed from the study of quantum information processing. The first and most obvious class contains quantum algorithms and protocols. The second is the use of reasoning about quantum systems to obtain insight into classical computer science. The third class consists of purely classical results inspired by the formalisms developed to deal with quantum information processing and quantum mechanics more generally. We briefly discuss this last class of applications, and then devote a section to each of the first two classes.

Researchers in quantum mechanics, responding to their need to delve deeply and carefully into the linear algebra and generalized probability theory underlying quantum mechanics, have developed powerful formalisms for discussing these areas. Dirac’s compact and suggestive bra/ket nota-

tion is useful for any work involving significant linear algebra. The operator view gives insight into classical probability theory, and understanding the tensor structure inherent in classical probability theory and its difference from a direct sum structure helps clarify many issues including relationships between joint distributions and their marginals.

Implications of reasoning about quantum systems to problems in classical computer science

We give two surprising, elegant examples.

Cryptographic protocols usually rely on the empirical hardness of a problem for their security; it is rare to be able to prove complete, information theoretic security. When a cryptographic protocol is designed based on a new problem, the difficulty of the problem must be established before the security of the protocol can be understood. Empirical testing of a problem takes a long time. Instead, whenever possible, “reduction” proofs are given that show that if the new problem were solved it would imply a solution to a known hard problem; the proofs show that the solution to the known problem can be reduced to a solution of the new problem. (Regev 2005) designed a novel, purely classical cryptographic system based on a certain problem. He was able to reduce a known hard problem to this problem, but only by using a quantum step as part of the reduction proof. Thus he has shown that if the new problem is efficiently solvable in any way, there is an efficient quantum algorithm for the old problem. But it says nothing about whether there would be a classical algorithm. This result is of practical importance; his new cryptographic algorithm is a more efficient lattice based public key encryption system. Lattice based systems are currently the leading candidate for public key systems secure against quantum attacks.

More spectacular, if less practical, is Aaronson’s new solution to a notorious conjecture involving a purely classical complexity class **PP** (Aaronson 2005b). From 1972 until 1995 this question remained open. Aaronson defines a new quantum complexity class **PostBQP**, an extension of the standard quantum complexity class **BQP**, motivated by the use of postselection in certain quantum arguments. It takes him a page to show that **PostBQP=PP**, and then only three lines to prove the conjecture. Thus it seems that for certain questions, the “right” way to view the classical class **PP** is through the eyes of quantum information processing.

Quantum algorithms and protocols

Shor’s factoring and discrete log algorithms solve important but narrow problems. Grover’s algorithm and its generalizations are applicable only to a more restricted class of problems than many people outside the field realize. For example, it is unfortunate that Grover used “database” in the title of (Grover 1997) since his algorithm does not apply to what most people mean by a database. Grover’s algorithm only gives a speed-up over unstructured search, and databases, which are generally highly structured, can be searched extremely rapidly classically. At best quantum computation can only give a constant factor improvement for searches of ordered data like that of databases (Childs, Landahl, & Parrilo 2006).

Even worse, obtaining output from Grover’s algorithm destroys the quantum superposition, and recreating the superposition is often linear in N which negates the $O(\sqrt{N})$ benefit of the search algorithm. For this reason Grover’s algorithm and its generalizations are only applicable to searches over data that has a sufficiently uniform and quick generating function which can be used to quickly compute the superposition.

Finding new quantum algorithms has been exceedingly slow going. Some more recent algorithms include (Hallgren 2002) for solving Pell’s equations, (Watrous 2001) for the group black box model, (van Dam, Hallgren, & Ip 2003) for the shifted Legendre symbol problem. The first two are closely related to Shor’s algorithm - they are in the class of hidden subgroup problems - and the third makes heavy use of Fourier transforms. In the past five years a new family of quantum algorithms has been discovered that uses techniques of quantum walks to solve a variety of problems, some related to graphs, others to matrix products or commutativity in groups (Childs *et al.* 2002; Magniez, Santha, & Szegedy 2005; Magniez & Nayak 2005; Buhrman & Špalek 2006; Krovi & Brun 2007).

For many years Shor’s algorithm and Grover’s algorithm were viewed as widely different algorithms. Quantum learning theory (Bshouty & Jackson 1999; Servedio 2001; Gortler & Servedio 2004; Hunziker *et al.* 2003; Atici & Servedio 2005) is closely tied to both. Quantum learning descends from computational learning theory, a subfield of artificial intelligence. Computational learning is concerned with concept learning. Common models include exact learning and probably approximately correct (PAC) learning. A concept is modeled by its membership as given by a Boolean function $c : \{0, 1\}^n \rightarrow \{0, 1\}$. Let $C = \{c_i\}$ be a concept class. Say one has access to an oracle O_c for one of the concepts c in C , but one doesn’t know which. The types of oracles assumed vary, but a common one is a membership oracle which upon input of x outputs $c(x)$. In the quantum case, one can input superpositions of inputs to obtain superpositions of outputs. One can ask a variety of questions as to how quickly and with how many queries to the oracle can the concept c be determined. Sample results in this area include the negative result that the number of classical and quantum queries required for any concept class does not differ by more than a polynomial in either the exact or PAC model. However the story is different if computational efficiency is taken into account. In the exact model the existence of *any* classical one-way function guarantees the existence of a concept class which is polynomial-time learnable in the quantum case but not in the classical. For the PAC model a slightly weaker result is known in terms of a particular one-way function.

Probability theory and quantum mechanics

To quote Scott Aaronson (Aaronson 2005a):

“To describe a state of n particles, we need to write down an exponentially long vector of exponentially small num-

bers, which themselves vary continuously. Moreover, the instant we measure a particle, we “collapse” the vector that describes its state - and not only that, but possibly the state of another particle on the opposite side of the universe. Quick, what theory have I just described?”

“The answer is classical probability theory. The moral is that, before we throw up our hands over the “extravagance” of the quantum worldview, we ought to ask: is it so much more extravagant than the classical probabilistic worldview? After all, both involve linear transformations of exponentially long vectors that are not directly observable.”

We spend the next section putting this view of probability theory on a firm basis. We then describe how quantum mechanics is a formal extension of probability theory. We only sketchily describe this extension; more details can be found in (Strocchi 2005; Kuperberg 2005; Redei & Summers 2006; Kitaev, Shen, & Vyalys 2002; Sudbery 1986; Mackey 1963).

Many, but not all, of the unintuitive aspects of quantum mechanics exist in classical probability theory. Entanglement does not exist in classical probability, but classical correlations are strange enough, judging by human reaction to many of them.

A view of classical probability theory

Let A be a set of n elements. A probability distribution μ on A is a function

$$\mu : A \rightarrow [0, 1]$$

such that $\sum_{a \in A} \mu(a) = 1$. The space \mathcal{P}^A of all probability distributions over A has dimension $n - 1$. We can view \mathcal{P}^A as the $n - 1$ dimensional simplex $\sigma_{n-1} = \{x \in \mathbf{R}^n | x_i \geq 0, x_1 + x_2 + \dots + x_n = 1\}$ which is contained in the n dimensional space \mathbf{R}^A , the space of all functions from A to \mathbf{R} ,

$$\mathbf{R}^A = \{f : A \rightarrow \mathbf{R}\}.$$

For $n = 2$, the simplex σ_{n-1} is the line segment from $(1, 0)$ to $(0, 1)$. The vertices of the simplex correspond to the elements $a \in A$: a probability distribution μ maps to the point in the simplex $x = (\mu(a_1), \mu(a_2), \dots, \mu(a_n))$.

Let B be a set of m elements. Let $A \times B$ be the Cartesian product $A \times B = \{(a, b) | a \in A, b \in B\}$. What is the relation between $\mathcal{P}^{A \times B}$, the space of all probability distributions over $A \times B$, and the spaces \mathcal{P}^A and \mathcal{P}^B ? The tempting guess is not correct: $\mathcal{P}^{A \times B} \neq \mathcal{P}^A \times \mathcal{P}^B$. We see this relation does not hold by checking dimensions. First consider the relationship between $\mathbf{R}^{A \times B}$ and \mathbf{R}^A and \mathbf{R}^B . Since $A \times B$ has cardinality $|A \times B| = |A||B| = nm$, $\mathbf{R}^{A \times B}$ has dimension nm , which is not equal to $n + m$, the dimension of $\mathbf{R}^A \times \mathbf{R}^B$. Since in general $\dim(\mathcal{P}^A) = \dim(\mathbf{R}^A) - 1$, $\dim(\mathcal{P}^{A \times B}) = nm - 1$ which is not equal to $n + m - 2$, the dimension of $\mathcal{P}^A \times \mathcal{P}^B$, so $\mathcal{P}^{A \times B} \neq \mathcal{P}^A \times \mathcal{P}^B$. Instead $\mathbf{R}^{A \times B}$ is the tensor product $\mathbf{R}^A \otimes \mathbf{R}^B$ of \mathbf{R}^A and \mathbf{R}^B . So $\mathcal{P}^{A \times B} \in \mathbf{R}^A \otimes \mathbf{R}^B$.

Tensor products are rarely mentioned in probability textbooks, but the tensor product is as much a part of probability theory as of quantum mechanics. The tensor product structure inherent in probability theory should be stressed more often; one of the sources of mistaken intuition about

probabilities is a tendency to try to impose the more familiar direct product structure on what is actually a tensor product structure. We briefly review tensor products here; readers not familiar with tensor products should consult more extensive expositions (Rieffel & Polak 2000; Nielsen & Chuang 2001; Rieffel & Polak in preparation).

The *tensor product* $V \otimes W$ of two vector spaces V and W with bases $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_m\}$ respectively is an nm -dimensional vector space with basis $a_i \otimes b_j$ where \otimes is the tensor product, an abstract binary operator defined by the following relations:

$$\begin{aligned} (v_1 + v_2) \otimes x &= v_1 \otimes x + v_2 \otimes x \\ v \otimes (x_1 + x_2) &= v \otimes x_1 + v \otimes x_2 \\ (\alpha v) \otimes x &= v \otimes (\alpha x) = \alpha v \otimes x. \end{aligned}$$

Taking $k = \min(n, m)$, all elements of $V \otimes X$ have form

$$v_1 \otimes w_1 + v_2 \otimes w_2 + \dots + v_k \otimes w_k.$$

Due to the relations defining the tensor product such a representation is not unique. Furthermore, most elements of $V \otimes W$ cannot be written as $v \otimes w$ where $v \in V$ and $w \in W$.

Let $A_0 = \{0_0, 1_0\}$, $A_1 = \{0_1, 1_1\}$, and $A_2 = \{0_2, 1_2\}$, where 1_0 versus 0_0 corresponds to whether or not the next person you meet is interested in quantum mechanics, A_1 to whether they know the solution to the Monty Hall problem, and A_2 to whether they are at least 5'6" tall. So $1_0 1_0 0_0$ corresponds to someone under 5'6" who is interested in quantum mechanics and knows the solution to the Monty Hall problem. We often write 110 instead of $1_0 1_0 0_0$; the subscripts are implied by the position. A probability distribution over the set of eight possibilities, $A_0 \times A_1 \times A_2$, has form

$$\vec{p} = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}).$$

More generally, a probability distribution over $A_0 \times A_1 \times \dots \times A_k$, where the A_i are all 2 element sets, is a vector of length 2^k . We now understand the first part of Aaronson's remark: vectors in probability theory are exponentially long.

Given functions $f : A \rightarrow \mathbf{R}$ and $g : B \rightarrow \mathbf{R}$, define the tensor product $f \otimes g : A \times B \rightarrow \mathbf{R}$ by $(a, b) \mapsto f(a)g(b)$. If μ and ν are probability distributions, then so is $\mu \otimes \nu$. The linear combination of distributions is a distribution as long as the linear coefficients are non-negative and sum to 1. Conversely, any distribution $\eta \in \mathcal{P}^{A \times B}$ is a linear combination of distributions of the form $\mu \otimes \nu$ with linear factors summing to 1.

A joint distribution $\mu \in \mathcal{P}^{A \times B}$ is *independent* or *uncorrelated* if it can be written as a tensor product $\mu_A \otimes \mu_B$ of distributions $\mu_A \in \mathcal{P}^A$ and $\mu_B \in \mathcal{P}^B$. The vast majority of joint distributions do not have this form, in which case they are *correlated*. For any joint distribution $\mu \in \mathcal{P}^{A \times B}$, we can define a *marginal* distribution $\mu_A \in \mathcal{P}^A$ by

$$\mu_A : a \mapsto \sum_{b \in B} \mu(a, b).$$

An uncorrelated distribution is the tensor product of its marginals. Other distributions cannot be reconstructed from their marginals; information has been lost.

A distribution μ on a finite set A that is concentrated entirely at one element is said to be a *pure*; on a set A of n elements there are exactly n pure distributions $\mu_a : A \rightarrow [0, 1]$, one for each element of A , where

$$\mu_a : a' \mapsto \begin{cases} 1 & \text{if } a' = a \\ 0 & \text{otherwise.} \end{cases}$$

All other distributions are said to be *mixed*.

Let us return to the example of the traits for the next person you meet. Unless you know all of these traits, the distribution $\vec{p} = (p_{000}, \dots, p_{111})$ is a mixed distribution. When you meet the person you can observe their traits. Once you have made these observations, the distribution “collapses” to a pure distribution. For example, if the person is interested in quantum mechanics, does not know the solution to the Monty Hall problem, and is 5’8”, the “collapsed” distribution is $\vec{p}_c = (0, 0, 0, 0, 0, 1, 0, 0)$.

To understand the final part of Aaronson’s remark, consider another example. Say someone prepares two sealed envelopes with identical pieces of paper and sends them to opposite sides of the universe. Half the time both envelopes contain 0; half the time 1. The initial distribution is $\vec{p}_I = (1/2, 0, 0, 1/2)$. If someone then opens one of the envelopes and observes a 0, the state of the contents of the other envelope is immediately known - known faster than light can travel between the envelopes - and the distribution “collapses” to $\vec{p}_u = (1, 0, 0, 0)$.

Are we disturbed by the “extravagance” of the exponential state space of classical probability theory, and the “faster-than-light collapse” of these classical vectors under observation? Another question one might ask is: can this “extravagance” be used to facilitate computation? The answer is a resounding yes; allowing randomness does give additional computational power. See (Harel 1987; Traub & Werschulz 1999) for delightful expositions of the computational benefits of randomness.

To fully understand the relationship between quantum mechanics and probability theory it is useful to view probability distributions as operators. Consider the set of linear operators $\mathcal{M}^A = \{M : \mathbf{R}^A \rightarrow \mathbf{R}^A\}$. To every function $f : A \rightarrow \mathbf{R}$, there is an associated operator $M_f : \mathbf{R}^A \rightarrow \mathbf{R}^A$ given by $M_f : g \mapsto fg$. An operator M is said to be a projector if $M^2 = M$. The probability distributions μ whose corresponding operators M_μ are projectors are exactly the pure distributions. The matrix for the operator corresponding to a function is always diagonal; for a probability distribution, diagonal and trace 1. For example, the operator corresponding to the probability distribution $\vec{p}_I = (1/2, 0, 0, 1/2)$ is represented by the matrix

$$\begin{pmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}.$$

Quantum mechanics as a generalization of probability theory

The vector representation of a quantum state has redundancy that can be confusing; any vector multiplied by a unit length

complex number $e^{i\theta}$ - called the global phase - represents the same quantum state. Another way of representing quantum states removes this ambiguity and makes the relation with probability theory clearer. We follow Dirac’s elegant and compact bra/ket notation. The row vector $\langle v|$ is the conjugate transpose of the column vector $|v\rangle$. For any N dimensional vector $|v\rangle$ representing a quantum state we can construct a density operator, the $N \times N$ matrix $|v\rangle\langle v|$. The density operator $|v\rangle\langle v|$ representing a quantum state no longer has ambiguity due to the global phase. Like the operators corresponding to probability distributions, the operators corresponding to quantum states have trace 1 and are positive and Hermitian. Density operators corresponding to quantum states $|v\rangle$ are projectors so have rank 1. Unlike operators for probability distributions, density operators need not be diagonal. For example, the density operator for the state $|\nearrow\rangle = 1/\sqrt{2}(|0\rangle + |1\rangle)$ is

$$|\nearrow\rangle\langle\nearrow| = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

This example illustrates that superpositions are distinct from mixtures of basis states since such mixtures must be diagonal: the fifty-fifty mixture of $|0\rangle$ and $|1\rangle$ has density operator

$$\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

The analog of taking the marginal is taking the partial trace. The partial trace $\text{tr}_W O_{VW}$ of an operator $\rho : V \otimes W \rightarrow V \otimes W$ with respect to the subsystem W is the operator

$$\rho_V = \text{tr}_W O_{VW} = \sum_i \langle b_i | O_{VW} | b_i \rangle$$

that acts on subsystem V , where $\{|b_i\rangle\}$ is a orthonormal basis for W . Taking the partial trace of a density operator produces another density operator, a Hermitian, positive, trace 1 operator. Density operators obtained from the partial trace model what can be learned about a subsystem from measurements on that subsystem alone. In this context they are often called mixed states. Density operators of the form $|v\rangle\langle v|$ are called pure states, or just quantum states. For example, the Bell state $|\Phi^+\rangle = 1/\sqrt{2}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) = 1/\sqrt{2}(|00\rangle + |11\rangle)$ has density operator

$$|\Phi^+\rangle\langle\Phi^+| = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

and its partial trace with respect to either one of its qubits is the 2-dim density operator $\frac{1}{2}I$.

Since every Hermitian operator can be diagonalized, every density operator ρ can be written as $\sum_i p_i |\psi_i\rangle\langle\psi_i|$, a probability distribution over pure quantum states where the $|\psi_i\rangle$ are mutually orthogonal eigenvectors of ρ , and p_i are the eigenvalues. Conversely any probability distribution μ over a set of orthogonal quantum states $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_L\rangle$ where $\mu : |\psi_i\rangle \rightarrow p_i$ has a corresponding density operator $\rho_\mu = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. In the basis $\{|\psi_i\rangle\}$, the density

operator ρ_μ is diagonal with entries p_1, \dots, p_L . Under the isomorphism between \mathbf{R}^L and the subspace of V generated by $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_L\rangle$, the density operator ρ_μ realizes the operator M_μ . Thus a probability distribution over a set of orthonormal quantum states $\{|\psi_i\rangle\}$ can be viewed as a trace 1 diagonal matrix acting on \mathbf{R}^L .

Although every density operator can be viewed as a probability distribution over a set of orthogonal quantum states, this representation is not in general unique. More importantly, for most pairs of density operators ρ_1 and ρ_2 , there is no basis over which both ρ_1 and ρ_2 are diagonal. In particular, only if ρ_1 and ρ_2 commute are they simultaneously diagonalizable, so only in this case can they both be viewed as probability distributions over the same set of states. Thus, although each density operator of dimension N can be viewed as a probability distribution over N states, the space of all density operators is much larger than the space of probability distributions over N states. Let $\rho : V \rightarrow V$ be a density operator. A density operator ρ corresponds to a pure state if and only if it is a projector. This statement is analogous to that for probability distributions; the pure states correspond exactly to rank 1 density operators, and mixed states have rank greater than 1. Density operators are also used to model probability distributions over pure states, particularly probability distributions over the possible outcomes of a measurement yet to be performed. Their use here is analogous to the classical use of probability distributions to model the probabilities of possible traits before they can be observed.

A pure quantum state $|\psi\rangle$ is *entangled* if it cannot be written as the tensor product of single qubit states. For a mixed quantum state, it is important to determine if all of its correlation comes from being a mixture in the classical sense or if it is also correlated in a quantum fashion. A mixed quantum state $\rho : V \otimes W \rightarrow V \otimes W$ is said to be *uncorrelated* if $\rho = \rho_V \otimes \rho_W$ for some density operators $\rho_V : V \rightarrow V$ and $\rho_W : W \rightarrow W$. Otherwise ρ is said to be *correlated*. A mixed quantum state ρ is said to be *separable* if it can be written $\rho = \sum_{j=1}^L p_j |\psi_j^V\rangle\langle\psi_j^V| \otimes |\phi_j^W\rangle\langle\phi_j^W|$ where $|\psi_j^V\rangle \in V$ and $|\phi_j^W\rangle \in W$. In other words, ρ is separable if all the correlation comes from its being a classical mixture of uncorrelated quantum states. If a mixed state ρ is not separable it is *entangled*. For example, the mixed state $\rho_{cc} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)$ is classically correlated but not entangled whereas the Bell state $|\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|)$ is entangled. The marginals of a pure distribution are always pure, but the analogous statement is not true for quantum states; all of the partial traces of a pure state are pure only if the original pure state was not entangled. As we saw, the partial traces of the Bell state $|\Phi^+\rangle$, a pure state, are not pure. Most pure quantum states are entangled, exhibiting quantum correlations with no classical analog. All pure probability distributions are completely uncorrelated.

Classical and quantum analogs:

Classical probability	Quantum mechanics
probability distribution μ viewed as operator M_μ	density operator ρ
pure dist: M_μ is a projector	pure state: ρ is a projector
marginal distribution	partial trace
A distribution is <i>uncorrelated</i> if it is the tensor product of its marginals	A state is <i>uncorrelated</i> if it is the tensor product of its partial traces
Key difference:	
Classical probability	Quantum mechanics
pure distributions are always uncorrelated	pure states contain no classical correlation but can be entangled

Where does the power of quantum information processing come from?

Quantum parallelism?

For any classical computation of a function $f(x)$ on n bits, the analogous quantum computation U_f produces a superposition $\frac{1}{\sqrt{N}} \sum |x, f(x)\rangle$ of all input/output pairs upon input of a superposition of all input values. The ability of a quantum computer to obtain a superposition of all input/output pairs with similar effort as it takes a classical computer to obtain a single pair is called *quantum parallelism*. Since quantum parallelism enables one to work simultaneously with 2^n values, it in some sense circumvents the time/space trade-off of classical parallelism through its ability to hold exponentially many computed values in a linear amount of physical space. However, this effect is less powerful than it may initially appear.

We can gain only limited information from this superposition: these 2^n values of f are not independently accessible. We only gain information by measuring, but measuring in the standard basis projects the final state onto a single input/output pair $|x, f(x)\rangle$, and a random one at that. By itself, quantum parallelism is useless.

While $N = 2^n$ output values of $f(x)$ appear in the single superposition state, it still takes $N = 2^n$ computations of U_f to obtain them all, no better than the classical case. This limitation leaves open the possibility that quantum parallelism can help in cases where only a single output, or a small number of outputs, is desired. It suggests an exponential speed up, but such speed ups are rare. It has been proven that no quantum algorithm can improve on the $O(\sqrt{N})$ that Grover's algorithm achieves for unstructured search (Bennett *et al.* 1997), and for many other problems it has been proven that quantum computation cannot provide any speed-up (Beals *et al.* 2001; Ambainis 2000).

Exponential size of quantum state space?

As we have seen, exponential spaces also arise in classical probability theory. Furthermore, what would it mean for an efficient algorithm to take advantage of the exponential size of a space? A superposition like $\frac{1}{\sqrt{N}} \sum |x, f(x)\rangle$ is only a single state of the quantum state space. The vast majority

of states cannot even be approximated by an efficient quantum algorithm (Knill 1995). An efficient quantum algorithm cannot even come close to most states in the state space. So quantum parallelism does not, and efficient quantum algorithms cannot, make use of the full state space.

Quantum Fourier transforms?

Most quantum algorithms use quantum Fourier transforms (QFTs). The Walsh-Hadamard transformation, a QFT over the group \mathbf{Z}_2 , is frequently used to create a superposition of 2^n input values. In addition the heart of most quantum algorithms makes use of QFTs. Shor and Grover use QFTs in both of these ways. Many researchers speculated that quantum Fourier transforms are the paramount quantum resource for quantum computation. So it came as a surprise when (Aharonov, Landau, & Makowsky 2006) showed that the QFT is classically simulatable. Given the ubiquity of quantum Fourier transforms in quantum algorithms, researchers continue to consider QFTs as one of the main tools of quantum computation, but in themselves they are not sufficient.

Entanglement?

(Jozsa & Linden 2003) show that any quantum algorithm involving only pure states that achieves exponential speed-up over classical algorithms must entangle a large number of qubits. While entanglement is necessary for an exponential speed-up, the existence of entanglement is far from sufficient to guarantee a speed-up, and it may turn out that another property better characterizes what gives a speed-up. Many entangled systems have been shown to be classically simulatable (Vidal 2003; Markov & Shi 2005). Furthermore, if one looks at query complexity instead of algorithmic complexity, an exponential benefit can be obtained without any entanglement whatsoever. (Meyer 2000) shows that in the course of the Bernstein-Vazirani algorithm, which achieves an N to 1 reduction in the number of queries required, no qubits become entangled. More obviously the BB84 quantum key distribution protocol makes no use of entanglement.

For these reasons entanglement should not be viewed as the sole source of power in quantum information processing. However it is important in many contexts, and required in others. While researchers have long recognized entanglement as a uniquely quantum resource, much about entanglement is poorly understood. Entanglement with respect to tensor decompositions of only two factors is completely characterized for pure states, and well studied for mixed states. See (Bruss 2002) for an introductory survey. But understanding bi-partite entanglement is of limited utility for understanding quantum computation because there we are interested in entanglement between large numbers of qubits. Full characterization of entanglement with respect to tensor decompositions with many factors is difficult; where in the bi- or tri-partite cases only a finite number of parameters are needed, infinitely many parameters are required for four or more tensor factors (Dür, Vidal, & Cirac 2000).

Instead of trying to fully characterize multipartite entanglement, we can ask which types of entanglement are

useful, and for what. Significant progress has been made here, though much work remains. Cluster states were discovered to be a universal resource for quantum computation. In cluster state, or one-way, quantum computing (Raussendorf, Browne, & Briegel 2003; Nielsen 2005) a highly entangled “cluster” states is set up at the beginning of the algorithm. All computations take place by single qubit measurements, so the entanglement between the qubits can only decrease in the course of the algorithm (the reason for the “one-way” name). The initial cluster state is independent of the algorithm to be performed; it depends only on the size of the problem to be solved. In this way cluster state quantum computation makes a clean separation between the entanglement creation and the computational stages. While the cluster state model clarifies somewhat the role of entanglement in quantum computation, in another model, adiabatic quantum computation (Aharonov *et al.* 2004), which like the cluster state model has been proved equivalent to the standard circuit model of quantum computation, the role of entanglement is obscure. Many intriguing questions as to the source of power in quantum information processing remain, and are likely to remain for many years while we humans struggle to understand what Nature allows us to compute quickly and why.

References

- [Aaronson 2005a] Aaronson, S. 2005a. Are quantum states exponentially long vectors? quant-ph/0507242.
- [Aaronson 2005b] Aaronson, S. 2005b. Quantum computing, postselection, and probabilistic polynomial-time. *Proceedings of the Royal Society A* 461:3473–3482.
- [Aharonov *et al.* 2004] Aharonov, D.; van Dam, W.; Kempe, J.; Landau, Z.; Lloyd, S.; and Regev, O. 2004. Adiabatic quantum computation is equivalent to standard quantum computation. [/quant-ph/0405098](http://quant-ph/0405098).
- [Aharonov, Landau, & Makowsky 2006] Aharonov, D.; Landau, Z.; and Makowsky, J. 2006. The quantum FFT can be classically simulated. Los Alamos Physics Preprint Archive, <http://xxx.lanl.gov/abs/quant-ph/0611156>.
- [Ambainis 2000] Ambainis, A. 2000. Quantum lower bounds by quantum arguments. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, 636–643.
- [Atici & Servedio 2005] Atici, A., and Servedio, R. 2005. Improved bounds on quantum learning algorithms. *Quantum Information Processing* 4(5):355–386.
- [Beals *et al.* 2001] Beals, R.; Buhrman, H.; Cleve, R.; Mosca, M.; and de Wolf, R. 2001. Quantum lower bounds by polynomials. *J. ACM* 48(4):778–797.
- [Bennett *et al.* 1997] Bennett, C. H.; Bernstein, E.; Brassard, G.; and Vazirani, U. V. 1997. Strengths and weaknesses of quantum computing. *Society for Industrial and Applied Mathematics Journal on Computing* 26(5):1510–1523.
- [Bennett 1989] Bennett, C. H. 1989. Time/space trade-offs

- for reversible computation. *Society for Industrial and Applied Mathematics Journal on Computing* 18(4):766–776.
- [Brassard, Høyer, & Tapp 1998] Brassard, G.; Høyer, P.; and Tapp, A. 1998. Quantum counting. *Lecture Notes in Computer Science* 1443:820–831.
- [Bruss 2002] Bruss, D. 2002. Characterizing entanglement. *Journal of Mathematical Physics* 43(9):4237–4250.
- [Bshouty & Jackson 1999] Bshouty, N. H., and Jackson, J. C. 1999. Learning DNF over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing* 28:1136–1142.
- [Buhrman & Špalek 2006] Buhrman, H., and Špalek, R. 2006. Quantum verification of matrix products. In *Proc. of 17th ACM-SIAM SODA*, 880–889.
- [Childs *et al.* 2002] Childs, A. M.; Cleve, R.; Deotto, E.; Farhi, E.; Gutmann, S.; and Spielman, D. A. 2002. Exponential algorithmic speedup by quantum walk. [quant-ph/0209131](#).
- [Childs, Landahl, & Parrilo 2006] Childs, A. M.; Landahl, A. J.; and Parrilo, B. A. 2006. Improved quantum algorithms for the ordered search problem via semidefinite programming. [quant-ph/0608161](#).
- [Cohen 1995] Cohen, L. 1995. *Time-frequency analysis*. Prentice Hall.
- [Deutsch & Jozsa 1992] Deutsch, D., and Jozsa, R. 1992. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London Ser. A* A439:553–558.
- [Deutsch 1985] Deutsch, D. 1985. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proceedings of the Royal Society of London Ser. A* A400:97–117.
- [Deutsch 1998] Deutsch, D. 1998. David Deutsch’s many worlds. *Frontiers*.
- [Dür, Vidal, & Cirac 2000] Dür, W.; Vidal, G.; and Cirac, J. I. 2000. Three qubits can be entangled in two inequivalent ways. *Physical Review A* 62.
- [Einstein, Podolsky, & Rosen 1935] Einstein, A.; Podolsky, B.; and Rosen, N. 1935. Can quantum-mechanical description of physical reality be considered complete? *Physical Review* 47:777–780.
- [Gortler & Servedio 2004] Gortler, S., and Servedio, R. 2004. Equivalences and separations between quantum and classical learnability. *SIAM Journal on Computing* 33(5):1067–1092.
- [Greenberger *et al.* 1990] Greenberger, D. M.; Horne, M. A.; Shimony, A.; and Zeilinger, A. 1990. A Bell’s theorem without inequalities. *American Journal of Physics* 58:1131–1143.
- [Greenberger, Horne, & Zeilinger 1989] Greenberger, D. M.; Horne, M. A.; and Zeilinger, A. 1989. Going beyond Bell’s theorem. In *Bell’s Theorem, Quantum Theory and Conceptions of the Universe*. Kluwer. 73–76.
- [Grover 1997] Grover, L. K. 1997. Quantum computers can search arbitrarily large databases by a single query. *Physical Review Letters* 79(23):4709–4712.
- [Hallgren 2002] Hallgren, S. 2002. The hidden subgroup problem and quantum computation using group representations. *STOC*.
- [Harel 1987] Harel, D. 1987. *Algorithmics: The spirit of computing*. Addison-Wesley.
- [Hunziker *et al.* 2003] Hunziker, M.; Meyer, D. A.; Park, J.; Pommersheim, J.; and Rothstein, M. 2003. The geometry of quantum learning. [quant-ph/0309059](#).
- [Jozsa & Linden 2003] Jozsa, R., and Linden, N. 2003. On the role of entanglement in quantum computational speedup. *Proceedings of the Royal Society of London Ser. A* 459:2011–2032.
- [Kitaev, Shen, & Vyalı 2002] Kitaev, A. Y.; Shen, A. H.; and Vyalı, M. N. 2002. *Classical and Quantum Computation*. American Mathematical Society.
- [Knill 1995] Knill, E. 1995. Approximation by quantum circuits. [quant-ph/9508006](#).
- [Krovi & Brun 2007] Krovi, H., and Brun, T. A. 2007. Quantum walks on quotient graphs. [quant-ph/0701173](#).
- [Kuperberg 2005] Kuperberg, G. 2005. A concise introduction to quantum probability, quantum mechanics, and quantum computation. unpublished notes.
- [Mackey 1963] Mackey, G. W. 1963. *Mathematical foundations of quantum mechanics*. W. A. Benjamin, Inc.
- [Magniez & Nayak 2005] Magniez, F., and Nayak, A. 2005. Quantum complexity of testing group commutativity. *ICALP*.
- [Magniez, Santha, & Szegedy 2005] Magniez, F.; Santha, M.; and Szegedy, M. 2005. Quantum algorithms for the triangle problem. *Proc. SODA*.
- [Markov & Shi 2005] Markov, I., and Shi, Y. 2005. Simulating quantum computation by contracting tensor networks. [quant-ph/0511069](#).
- [Meyer 2000] Meyer, D. A. 2000. Sophisticated quantum search without entanglement. *Physical Review Letters* 85:2014–2017.
- [Nielsen & Chuang 2001] Nielsen, M., and Chuang, I. 2001. *Quantum Computing and Quantum Information*. Cambridge Press.
- [Nielsen 2005] Nielsen, M. A. 2005. Cluster-state quantum computation. [quant-ph/0504097](#).
- [Pan *et al.* 2000] Pan, J.; Bouwmeester, D.; Daniell, M.; Weinfurter, H.; and Zeilinger, A. 2000. Experimental test of quantum nonlocality in three-photon Greenberger-Horne-Zeilinger entanglement. *Nature* 403:515–518.
- [Raussendorf, Browne, & Briegel 2003] Raussendorf, R.; Browne, D. E.; and Briegel, H. J. 2003. Measurement-based quantum computation with cluster states. *Physical Review A* 68:022312.
- [Redei & Summers 2006] Redei, M., and Summers, S. J. 2006. Quantum probability theory. [/hep-th/0601158](#).

- [Regev 2005] Regev, O. 2005. On lattices, learning with errors, random linear codes, and cryptography. In *Proc. 37th ACM Symp. on Theory of Computing (STOC)*, 84–93.
- [Rieffel & Polak 2000] Rieffel, E. G., and Polak, W. 2000. An introduction to quantum computing for non-physicists. *ACM Computing Surveys* 32(3):300 – 335.
- [Rieffel & Polak in preparation] Rieffel, E. G., and Polak, W. in preparation. *A Gentle Introduction to Quantum Computing*. MIT Press.
- [Servedio 2001] Servedio, R. A. 2001. Separating quantum and classical learning. *Lecture Notes in Computer Science* 2076:1065–1080.
- [Strocchi 2005] Strocchi, F. 2005. *An Introduction to the Mathematical Structure of Quantum Mechanics*. World Scientific.
- [Sudbery 1986] Sudbery, A. 1986. *Quantum Mechanics and the Particles of Nature*. Cambridge University Press.
- [Traub & Werschulz 1999] Traub, J., and Werschulz, A. G. 1999. *Complexity and Information*. Cambridge University Press.
- [van Dam, Hallgren, & Ip 2003] van Dam, W.; Hallgren, S.; and Ip, L. 2003. Quantum algorithms for some hidden shift problems. In *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, 489–498.
- [Vidal 2003] Vidal, G. 2003. Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters* 91:147902.
- [Watrous 2001] Watrous, J. 2001. Quantum algorithms for solvable groups. *STOC* 60 – 67.